

Data Management, Analysis, and Interpretation

Masoud Kayhanian, Ph.D.

Senior Development Engineer

Center for Environmental and Water Resources Engineering

Department of Civil and Environmental Engineering

UC Davis

Previous Presentations

- Set up a monitoring program
- Collect representative samples
- Select correct analytical methods
- Validate laboratory and field data

Topics Covered in this Presentation

- Storing the monitoring data
- Retrieving the data
- Analyzing the data
- Interpreting the data

Important Considerations in Data Management

- Consistent data reporting
- Data analysis methods

Consistent Data Reporting

Caltrans reports monitoring data under the general categories of:

- Sample description
- Site description
- Event description
- Standardized constituent names

Examples of Sample Description

- Site ID
- Event ID
- Sample start and end date
- Sample source
- Constituent type
- Analytical method
- Reported value
- Numerical qualifier
- Collection meetod

Examples of Sampling Event Description

- Event rain
- Max intensity
- Antecedent dry period
- Antecedent event rain
- Total flow volume
- Peak flow
- Percent capture
- Cumulative precipitation

Examples of Site Description

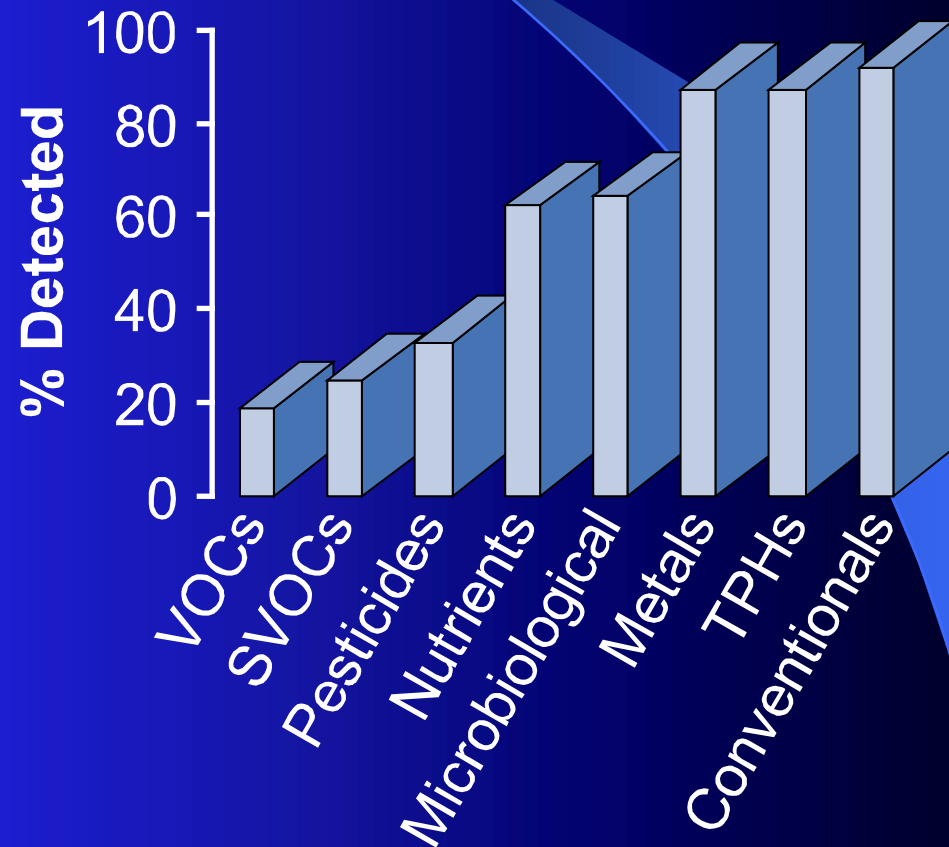
- Site description
- Caltrans district
- Hydrologic sub-area
- Land use
- Total flow volume
- Catchment area
- Impervious fraction
- BMP type
- Latitude/longitude
- Receiving water type
- Annual average daily traffic (AADT)

Examples of Standard Constituent Names

- $\text{NH}_3\text{-N}$
- Cr (III)
- Cr (VI)
- Fecal Coliform, Fecal Enterococci, Fecal Streptococci
- 1,1,1,2-Tetrachloroethane
- 1,1,2,2-Tetrachloroethane
- Oil and Grease
- TRPH, TVPH, TPH (diesel, gasoline, heavy oil, jet fuel, kerosene, motor oil)

Data Analysis: Addressing the Problem of Non-Detects

A large fraction of storm water quality data falls below detection limits (DL) and is reported as non-detect (ND).



Example Data Set: Dissolved Nickel, DL = 1

ND	11	5	10
5	4	4	ND
7	ND	10	4
15	7	ND	7
3	2	7	2
5	ND	11	ND
ND	3	ND	3
10	ND	ND	ND

Conventional Methods to Compute Mean (Non-Science Based)

- Ignore NDs (as if they do not exist!)
- $ND = 0$
- $ND = \text{Detection Limit}$
- $ND = \frac{1}{2} (\text{Detection Limit})$

Mean Values Based on Conventional Methods

Method	Mean Value
Ignore NDs	6.4
ND = 0	4.2
ND = DL	4.6
ND = $\frac{1}{2}$ (DL)	4.4

Statistical Approach to Estimating Mean (Science Based)

- **Cohen's Maximum Likelihood Estimation**
- **Maximum Likelihood Estimation (MLE) by Delta and Bootstrap Methods**
- **Regression on Order Statistics (ROS)**
- **EPA Delta Lognormal Statistical Method**

Mean Values For Selected Metal Constituents Based on Different Statistical Approach

Constituent	n	%ND	ROS	MLE	Cohen	EPA Delta Log
Al-D	25	20	187	139	106	141
As-T	46	28	4.1	4	3.2	4.2
Cd-T	373	31	1.3	1.3	0.9	1.3
Cr-D	383	5	11.3	11.3	11.0	11.1
Cr-D	462	46	2.1	2.5	2.8	1.1
Ni-D	481	39	4.0	4.1	2.6	4.1
Pb-D	523	31	5.9	5.9	1.75	5.3

Influence of Detection Limit

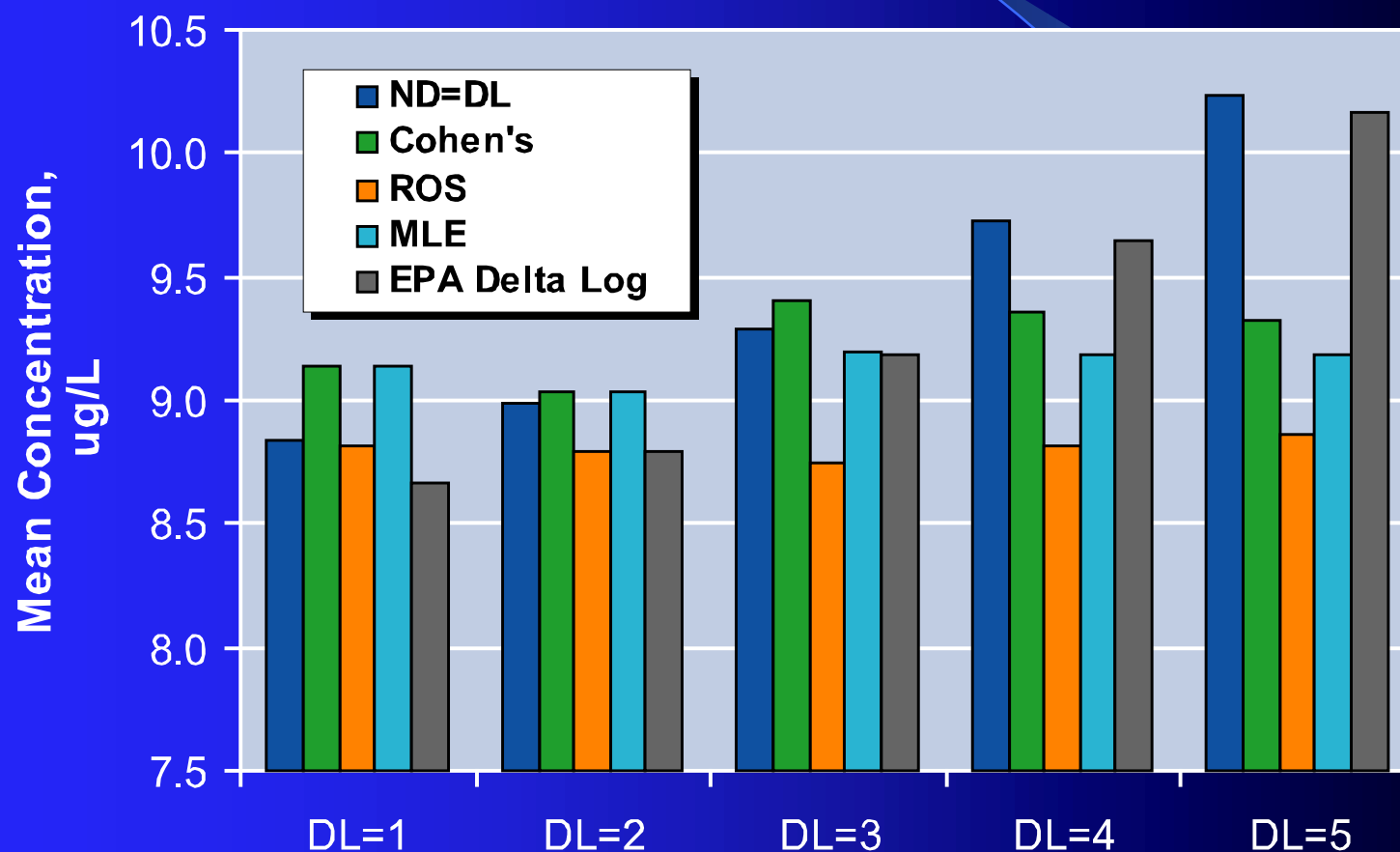
Lower Detection Limit

- Smaller numbers of non-detects
- Less need of science-based statistical approach
- Lower variability in mean computation
- Higher analytical cost

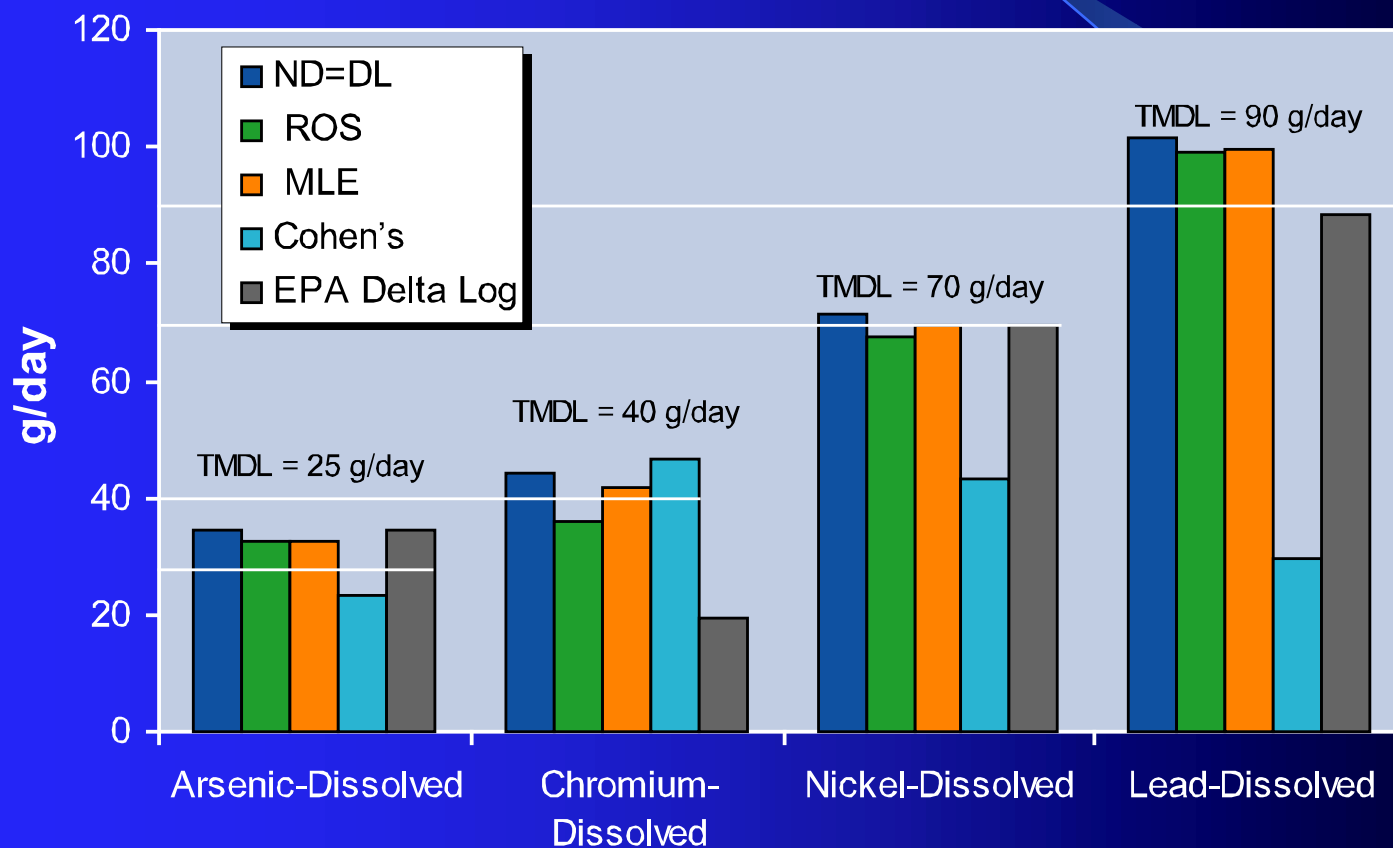
Higher Detection Limit

- Larger numbers of non-detects
- More need for science-based statistical approach
- Higher variability in mean calculation
- Lower analytical cost

Influence of Detection Limit on Mean Concentration



Implications for Total Maximum Daily Load (TMDL)



Caltrans Database Demonstration

Conclusions

- Without proper data QA/QC, the validity of monitoring data is questionable.
- All monitoring data should be reported in a consistent manner, in order to be fully utilized.
- A user-friendly database is essential to store and retrieve monitoring data for data analysis and to measure program success.
- A significant proportion of storm water runoff quality data contain large numbers of non-detects

Conclusions (continued)

- Detection limits set by analytical laboratories can affect the number of non-detects in water quality data.
- A large variation in calculated mean values can be observed depending on data distribution, number of NDs, and statistical method.
- Variation in calculated mean values can significantly affect the constituent mass loading estimation.
- Statistical approaches used in analyzing water quality data with non-detects may affect the TMDL compliance requirements.

Acknowledgements

- Camp Dresser and McKee (CDM)
- Civil and Environmental Engineering, UCD
- Civil and Environmental Engineering, UCLA
- Division of Environmental Analysis, Caltrans
- Geomatrix Consultants
- Kinnetic Laboratory Inc.
- Laboratory Data Consultants (LDC)
- Larry Walker Associates
- Law Crandall Engineering and Environmental Services
- Office of Water Program, CSU, Sacramento
- RFF Consulting
- URS Corporation